

# QA of AI

## Research Prompting Session with ChatGPT

By Michael S. Bridges

July 1, 2024

### Table of Contents

MB Q: Who is auditing and verifying that the AIs are behaving within the boundary constraints that have been set for it? How is QA performed on AI? .....	4
Internal Teams and Developers .....	4
Ethics Committees and Review Boards.....	4
External Auditors and Third-Party Assessors .....	4
Regulatory Bodies .....	4
Quality Assurance (QA) Teams.....	4
Continuous Monitoring and Feedback Loops .....	4
User Feedback and Transparency.....	4
Collaboration with External Organizations .....	5
Internationalized AI Models.....	5
Language Nuances .....	5
Cultural Norms and Sensitivities.....	5
Legal and Ethical Considerations .....	5
User Expectations and Engagement .....	5
Bias and Fairness.....	6
Localized Content and Information .....	6
User Interface and Design.....	6
Automated Testing of AI NLP Model .....	6
Test Automation for Scale.....	6
Creation of Automated Test Scripts.....	6
Test AI as a Test Oracle .....	7
Language Coverage .....	7
Handling Large Datasets .....	7
Regression Testing .....	7
Scalable Infrastructure.....	7

Dynamic Scenario Testing .....	7
Careers in AI QA Test .....	7
AI QA Engineer .....	8
NLP QA Tester .....	8
Machine Learning Tester .....	8
Data Quality Analyst (AI/ML) .....	8
Automation Test Engineer (AI).....	8
Conversational AI Tester .....	8
Ethical AI Tester .....	8
AI Test Architect.....	8
AI Test Certifications .....	9
ISTQB Advanced Level - Test Automation Engineer .....	9
Google Cloud Professional Machine Learning Engineer .....	9
Certified AI & ML Black Belt (Experfy) .....	9
Certified Software Test Professional (CSTP) - Machine Learning Specialization (iSQI) .....	9
Coursera Machine Learning Testing (University of Minnesota) .....	9
ML Tools.....	9
AWS SageMaker (Amazon Web Services).....	10
Google Cloud AI Platform (Google Cloud) .....	10
IBM Watson Studio (IBM Cloud) .....	10
Databricks Unified Analytics Platform .....	10
DataRobot .....	10
Alteryx Designer .....	10
ML Tools wrt Open-Source .....	10
AWS SageMaker:.....	10
Google Cloud AI Platform .....	10
IBM Watson Studio .....	11
Databricks Unified Analytics Platform .....	11
DataRobot .....	11
Alteryx Designer .....	11
Scikit-learn .....	11
TensorFlow.....	11
Apache Spark .....	11

Apache Spark Extension Capabilities .....	11
ML Glossaries .....	12
Google's Machine Learning Glossary .....	12
TensorFlow Glossary .....	12
Scikit-learn Glossary .....	12
Microsoft Azure Machine Learning Glossary .....	12
Data Science Glossary on Kaggle .....	12
Machine Learning Glossary on Towards Data Science .....	13
ML Overview .....	13
Initialize Weights and Bias .....	13
For Each Training Example .....	13
Update Weights .....	13
Repeat .....	13
ML Community Websites .....	13
Stack Overflow - Machine Learning .....	14
GitHub .....	14
Kaggle .....	14
Towards Data Science on Medium .....	14
ArXiv - Machine Learning Section .....	14
Reddit - r/MachineLearning .....	14
TensorFlow Forum .....	14
PyTorch Discussion Forum .....	14
Professional Associations for QA testing AI / ML Generative Models .....	14
International Software Testing Qualifications Board (ISTQB) .....	14
Association for Advancement of Artificial Intelligence (AAAI) .....	15
IEEE Standards Association .....	15
Guide to AI Tools .....	16
AI News .....	16

## MB Q: Who is auditing and verifying that the AIs are behaving within the boundary constraints that have been set for it? How is QA performed on AI?

ChatGPT A: Auditing and verifying the behavior of AI systems, especially in the context of adherence to ethical, legal, and safety constraints, is a complex process that involves interdisciplinary efforts. Different stakeholders play roles in ensuring the responsible development and deployment of AI. Here are some key aspects:

### Internal Teams and Developers

The development teams and AI researchers are responsible for implementing ethical principles and constraints during the design and coding phases. They must adhere to best practices, guidelines, and standards. Internal auditing and code reviews can be conducted to assess compliance.

### Ethics Committees and Review Boards

Organizations may establish ethics committees or review boards that include experts from diverse fields, such as AI ethics, philosophy, law, and sociology. These committees provide oversight and guidance to ensure that AI systems align with ethical and legal norms.

### External Auditors and Third-Party Assessors

Independent auditing firms or third-party assessors can be employed to conduct external audits of AI systems. These entities review the AI's behavior, assess its impact on users and society, and verify compliance with established ethical and regulatory frameworks.

### Regulatory Bodies

Governmental or industry-specific regulatory bodies may play a role in overseeing AI systems, particularly in sectors where strict regulations exist. These bodies define standards, conduct inspections, and ensure that AI systems comply with legal requirements.

### Quality Assurance (QA) Teams

QA teams, similar to those in traditional software development, can test AI systems to ensure they meet functional requirements and behave as expected. This may involve creating test scenarios, evaluating system outputs, and assessing performance under various conditions.

### Continuous Monitoring and Feedback Loops

Establishing continuous monitoring mechanisms and feedback loops is essential. AI systems can evolve and adapt over time, and ongoing monitoring helps detect and address any unintended consequences or ethical concerns that may arise during deployment.

### User Feedback and Transparency

Gathering feedback from end-users can provide valuable insights into the real-world impact of AI systems. Transparency in the decision-making process of AI algorithms is crucial, allowing users to understand and challenge the system's behavior.

## Collaboration with External Organizations

Organizations may collaborate with external entities, such as research institutions, non-governmental organizations (NGOs), and advocacy groups, to obtain independent assessments of their AI systems.

QA for AI involves a combination of technical, ethical, and societal considerations. It requires a holistic approach that goes beyond traditional software testing to encompass ethical considerations, bias detection, fairness assessments, and impact evaluations on various stakeholders.

As AI technology evolves, the development of standardized frameworks for auditing and verifying AI behavior is an ongoing area of research and industry collaboration. The goal is to ensure that AI systems are developed and deployed responsibly, taking into account the broader societal implications of their use.

## Internationalized AI Models

MB Q: do nuanced language and cultural differences? For example would the differences between the US vs France language and cultural norms, ethics, laws, etc. warrant a full retest of a multi-lingual AI's NLP and conversational model as though it were an entirely different system?

ChatGPT A: Nuanced language and cultural differences between regions like the US and France can indeed warrant a thorough retest of a multi-lingual AI's NLP (Natural Language Processing) and conversational model. Language and cultural nuances can significantly impact the effectiveness, appropriateness, and user experience of an AI system, especially when it interacts with users from different regions. Here are some factors to consider:

### Language Nuances

Nuances in language, including idioms, colloquial expressions, and linguistic variations, can differ between regions. A phrase that is perfectly acceptable in one region might have a different interpretation or connotation in another. Retesting ensures that the model understands and responds appropriately to language nuances specific to each target region.

### Cultural Norms and Sensitivities

Cultural norms and sensitivities vary widely between different countries and regions. Certain topics, jokes, or references that are acceptable in one culture may be inappropriate or offensive in another. Retesting helps identify and address potential cultural sensitivity issues, ensuring that the AI system respects local cultural norms.

### Legal and Ethical Considerations

Laws and ethical standards can differ significantly between countries. Retesting is crucial to ensure compliance with local regulations, privacy laws, and ethical guidelines specific to each region. This includes considerations related to data protection, user consent, and other legal requirements.

### User Expectations and Engagement

User expectations and engagement patterns can vary based on cultural preferences. Retesting helps assess how well the AI system engages users from different regions, providing a positive and culturally relevant experience. This may involve adjustments to the conversational style, tone, and content.

## Bias and Fairness

Language models are susceptible to bias, and biases can be introduced based on the data used for training. Retesting helps identify and mitigate biases that may disproportionately affect users from specific cultural or linguistic backgrounds, ensuring fairness in the AI system's responses.

## Localized Content and Information

Content and information provided by the AI may need to be localized for different regions. This involves adapting responses to include region-specific details, references, or recommendations. Retesting helps validate the accuracy and appropriateness of localized content.

## User Interface and Design

The user interface and design of the AI system may need to be adapted to accommodate language differences, text direction, and other cultural design preferences. Retesting ensures that the user interface is user-friendly and culturally appropriate for each target region.

While it may not always be necessary to treat language and cultural differences as entirely separate systems, a comprehensive retest allows for adjustments and optimizations tailored to each specific region. It's essential to take a user-centric approach, considering the diverse needs and expectations of users across different cultural and linguistic contexts. Additionally, ongoing monitoring and user feedback can contribute to continuous improvements in the AI system's performance across diverse regions.

## Automated Testing of AI NLP Model

MB Q: In the context of a QA team testing an AI system's NLP model, would the immense scope dictate the use of automated test techniques that can also scale and therefore inevitably arrive at the approach where by a test-AI is employed to automatically verify the AI system under test?

ChatGPT A: The immense scope of testing an AI system's NLP (Natural Language Processing) model, especially in a multi-lingual and culturally diverse context, often necessitates the use of automated testing techniques. Automated testing can help in dealing with the vast amount of data and scenarios that need to be covered. Employing a test AI (sometimes referred to as a test oracle) to automatically verify the AI system under test is indeed a viable approach. Here are some considerations:

### Test Automation for Scale

Automated testing is essential for scalability, enabling the QA team to cover a large number of test cases, variations, and potential scenarios efficiently. Given the immense scope, manual testing alone may be impractical.

### Creation of Automated Test Scripts

Develop automated test scripts that include a diverse set of test cases representing different languages, cultural contexts, and user interactions. These scripts should cover a wide range of inputs to assess the NLP model's performance under various conditions.

## Test AI as a Test Oracle

A test AI, acting as a test oracle, can be employed to automatically evaluate the correctness of the AI system's responses. The test AI can simulate user inputs, analyze model outputs, and compare them against expected results, helping to identify deviations or errors.

## Language Coverage

Use automated techniques to ensure comprehensive language coverage. Automated tests can be designed to handle multiple languages, dialects, and language variations, allowing the QA team to assess the NLP model's performance across diverse linguistic contexts.

## Handling Large Datasets

Automated testing is well-suited for handling large datasets, which is often the case in NLP testing. The test scripts can process and evaluate a substantial amount of textual data, enabling the QA team to assess the model's generalization capabilities.

## Regression Testing

Automated tests are valuable for conducting regression testing, ensuring that changes or updates to the AI system do not introduce new issues or negatively impact existing functionalities. This is crucial for maintaining the reliability and stability of the system.

## Scalable Infrastructure

Establish a scalable infrastructure for automated testing, allowing parallel execution of tests and efficient resource utilization. This helps in reducing the overall testing time and supports the frequent testing required in agile or continuous integration environments.

## Dynamic Scenario Testing

Automate the testing of dynamic scenarios that involve changing user inputs, evolving language patterns, and shifting contextual information. Automated tests can adapt to dynamic scenarios, providing a more realistic assessment of the NLP model's performance in real-world conditions.

It's important to note that while automated testing is valuable for scalability and efficiency, there are certain aspects of NLP testing, such as assessing subjective qualities like naturalness of language that may still benefit from manual testing and human evaluation.

In summary, a well-designed automated testing strategy, potentially leveraging a test AI as a test oracle, is a practical and efficient approach to address the immense scope of testing an AI system's NLP model, especially in a diverse and multi-lingual context.

## Careers in AI QA Test

MB Q: Regarding the QA testing of AI, have any specific QA testing job titles been created to recruit technical people?

ChatGPT A: Yes, as the field of QA testing for AI continues to evolve, specific job titles have emerged to reflect the unique skill set and expertise required for testing artificial intelligence systems. Some of these job titles include:

### AI QA Engineer

An AI QA Engineer is responsible for testing and ensuring the quality of AI-driven applications. This role involves creating test strategies, developing test cases, and evaluating the performance of AI algorithms.

### NLP QA Tester

Natural Language Processing (NLP) QA Testers specialize in testing AI systems that involve language understanding and generation. They assess how well NLP models comprehend and respond to diverse inputs, considering language nuances, semantics, and context.

### Machine Learning Tester

**Machine Learning Testers** focus on testing the machine learning components of AI systems. They evaluate the accuracy and generalization capabilities of machine learning models, identify biases, and assess the robustness of algorithms.

### Data Quality Analyst (AI/ML)

Data Quality Analysts in the context of AI and machine learning focus on ensuring the quality and integrity of the data used for training and testing AI models. They assess data quality, identify potential biases, and contribute to the creation of diverse and representative datasets.

### Automation Test Engineer (AI)

**Automation Test Engineers** specializing in AI focus on developing automated testing frameworks and scripts tailored for AI systems. They leverage automation tools to efficiently test various aspects of AI applications, including model outputs and system behavior.

### Conversational AI Tester

Conversational AI Testers specialize in evaluating the performance of conversational agents, chatbots, and virtual assistants. They assess how well these systems understand user queries, generate responses, and maintain context in dynamic conversations.

### Ethical AI Tester

Ethical AI Testers focus on evaluating AI systems from an ethical standpoint. They assess whether AI models exhibit biases, adhere to ethical guidelines, and comply with regulatory requirements. This role involves considering the ethical implications of AI decisions.

### AI Test Architect

AI Test Architects play a strategic role in designing test strategies and frameworks for AI applications. They contribute to the overall testing architecture, scalability, and efficiency of testing processes for complex AI systems.

These job titles reflect the growing need for specialized skills in testing AI technologies. Candidates for these roles typically require a combination of traditional QA testing expertise, domain knowledge in AI and machine learning, programming skills, and an understanding of the unique challenges posed by artificial intelligence systems. The exact title may vary across organizations, but the key focus is on testing and ensuring the quality of AI-driven applications.



## AI Test Certifications

MB Q: Regarding role of 'Machine Learning Tester', is there a industry-recognized certification?

ChatGPT A: As of my last knowledge update in January 2022, there isn't a specific industry-standard or universally recognized certification explicitly named "Machine Learning Tester." However, there are certifications and training programs that cover topics relevant to testing machine learning systems and AI applications. These certifications may be offered by various organizations, educational institutions, or certification bodies. It's important to note that the availability of certifications may have changed, and new certifications may have been introduced since my last update.

Here are some certifications and programs that are relevant to testing in the context of machine learning and AI:

### ISTQB Advanced Level - Test Automation Engineer

While not specific to machine learning, the ISTQB (International Software Testing Qualifications Board) Advanced Level - Test Automation Engineer certification covers advanced testing techniques and methodologies. It provides a foundation for understanding test automation, which is relevant when testing AI systems.

### Google Cloud Professional Machine Learning Engineer

Google Cloud offers a Professional Machine Learning Engineer certification, which covers designing, building, and deploying scalable machine learning models. While it focuses on engineering aspects, understanding machine learning practices is beneficial for testing.

### Certified AI & ML Black Belt (Experfy)

Experfy offers the Certified AI & ML Black Belt program, covering various aspects of AI and machine learning. While it is not specifically targeted at testing, gaining a deep understanding of AI concepts can be beneficial for testing roles.

### Certified Software Test Professional (CSTP) - Machine Learning Specialization (iSQI)

The International Software Quality Institute (iSQI) offers the CSTP with a Machine Learning Specialization. This program covers software testing principles with a focus on machine learning.

### Coursera Machine Learning Testing (University of Minnesota)

Some universities offer online courses and certifications related to machine learning testing. The University of Minnesota, for example, has a course on Coursera titled "Machine Learning Testing" that covers testing methodologies for machine learning systems.

When considering certifications, it's essential to evaluate the content, industry recognition, and relevance to your career goals. Additionally, staying informed about emerging certifications and training opportunities in the field of AI testing is advisable, as the landscape is dynamic and new programs may become available. Always check the official websites of certification providers for the latest information.

## ML Tools

MB Q: Are there other products similar to AzureML Studio?

ChatGPT A: Yes, there are several other platforms similar to AzureML Studio that provide tools for developing and deploying machine learning models. Some notable alternatives include

#### [AWS SageMaker \(Amazon Web Services\)](#)

A comprehensive machine learning platform on AWS offering end-to-end model development, training, and deployment with integrations to other AWS services.

#### [Google Cloud AI Platform \(Google Cloud\)](#)

Google Cloud's platform for building, deploying, and managing machine learning models, providing tools for data preparation, training, and serving models.

#### [IBM Watson Studio \(IBM Cloud\)](#)

IBM's platform for AI and machine learning that supports model development, experimentation, and deployment, with collaboration features and integration with various data sources.

#### [Databricks Unified Analytics Platform](#)

A cloud-based platform that integrates with popular data science and machine learning frameworks, offering collaborative data engineering and machine learning capabilities.

#### [DataRobot](#)

An automated machine learning platform that assists in building, deploying, and managing machine learning models without extensive coding.

#### [Alteryx Designer](#)

A data science and analytics platform that facilitates data preparation, blending, and model building with a focus on simplifying complex workflows.

These platforms provide various features and integrations tailored for different needs, and the choice depends on factors like cloud provider preference, specific use cases, and desired functionalities.

## [ML Tools wrt Open-Source](#)

MB Q: Are any of the 6 ML tools open-source?

ChatGPT A: Yes, some of the mentioned machine learning tools have open-source components or are built on open-source frameworks. Here's the breakdown:

#### [AWS SageMaker:](#)

While AWS SageMaker is a commercial product, it often integrates with open-source machine learning frameworks like TensorFlow and PyTorch. Users can bring their own open-source code and libraries for model development.

#### [Google Cloud AI Platform](#)

Google Cloud AI Platform supports popular open-source machine learning frameworks such as TensorFlow and scikit-learn. Users can bring their own code, and the platform provides integrations with these open-source tools.

## IBM Watson Studio

IBM Watson Studio integrates with various open-source tools and frameworks, allowing users to leverage popular libraries like TensorFlow, PyTorch, and scikit-learn. It supports collaboration on open-source Jupyter notebooks.

## Databricks Unified Analytics Platform

Databricks leverages Apache Spark, an open-source distributed computing system, as a core component. While Databricks itself is a commercial product, it incorporates and contributes to open-source projects.

## DataRobot

DataRobot is a commercial automated machine learning platform, and its core functionalities are not open source. However, it often supports the use of open-source models and frameworks.

## Alteryx Designer

Alteryx Designer is a commercial platform, and its core features are not open source. It focuses on providing a user-friendly interface for data science and analytics workflows.

It's worth noting that many machine learning tools and platforms integrate with open-source frameworks, allowing users to bring their own code, models, and libraries. Users can leverage open-source tools for specific tasks within these platforms. Additionally, there are standalone open-source tools and frameworks like **scikit-learn, TensorFlow, PyTorch, and Apache Spark** for users who prefer entirely open-source solutions.

## Scikit-learn

Scikit-learn is an open-source machine learning library for Python. It provides simple and efficient tools for data analysis and modeling, including classification, regression, clustering, and dimensionality reduction, making it a widely used resource for building and evaluating machine learning models.

## TensorFlow

TensorFlow is an open-source machine learning library developed by Google. It provides a comprehensive ecosystem for building and deploying machine learning models, emphasizing flexibility and scalability. TensorFlow is widely used for tasks such as deep learning, neural networks, and numerical computation.

## Apache Spark

Apache Spark is an open-source, distributed computing system for big data processing and analytics. It offers high-speed, fault-tolerant data processing using in-memory computation and supports various programming languages, making it widely used for large-scale data processing tasks.

## Apache Spark Extension Capabilities

Apache Spark has its own runtime environment and does not require Apache or Tomcat for its core functionality. However, Spark can be integrated with other Apache projects for extended capabilities. Some relevant dependencies include:

### *Hadoop Distributed File System (HDFS)*

While not a direct dependency, Spark can leverage HDFS for distributed storage when processing large datasets.

#### [Apache Hadoop YARN](#)

Spark can run on YARN, Hadoop's resource manager, to manage and allocate resources efficiently in a Hadoop cluster.

#### [Apache Hive and Apache HBase](#)

Spark can read data from Hive and HBase, integrating with these systems for data storage and retrieval.

#### [Apache Mesos](#)

Mesos can be used as an alternative to YARN for cluster management, allowing Spark to run on Mesos clusters.

#### [Apache Cassandra and Apache Kafka](#)

Spark can integrate with Cassandra and Kafka for distributed data storage and stream processing, respectively.

#### [S3 and other Cloud Storage](#)

Spark can read from and write to various cloud storage systems like Amazon S3, making it versatile for cloud-based data processing.

While Spark can function independently, its integration with other Apache projects and data storage systems enhances its capabilities for diverse big data processing scenarios.

## ML Glossaries

You can find online dictionaries and glossaries specific to machine learning (ML) vocabulary on various platforms. Some recommended sources include:

#### [Google's Machine Learning Glossary](#)

Google provides a comprehensive glossary of machine learning terms on its developers' website. You can find it by searching for "Google Machine Learning Glossary."

#### [TensorFlow Glossary](#)

TensorFlow, an open-source machine learning library, offers a glossary on its official documentation website. It covers terms related to TensorFlow and general machine learning concepts.

#### [Scikit-learn Glossary](#)

Scikit-learn, a popular machine learning library for Python, has a glossary section on its documentation website. It focuses on terms related to the library and machine learning in general.

#### [Microsoft Azure Machine Learning Glossary](#)

Microsoft Azure's documentation includes a machine learning glossary with terms related to Azure's machine learning services.

#### [Data Science Glossary on Kaggle](#)

Kaggle, a platform for data science and machine learning competitions, has a data science glossary that covers a wide range of terms used in the field.

## Machine Learning Glossary on Towards Data Science

Towards Data Science, a publication on Medium, has a comprehensive machine learning glossary that covers a broad range of terms and concepts.

When using online resources, it's important to refer to reputable sources, such as official documentation from machine learning libraries or platforms, to ensure accuracy and reliability. Additionally, community-driven platforms like Kaggle and Towards Data Science often provide valuable insights into practical applications and real-world examples of machine learning terminology.

## ML Overview

ChatGPT: Machine learning (ML) involves training algorithms on data to recognize patterns and make predictions or decisions without explicit programming. The process includes data collection, feature extraction, model training using labeled examples, and evaluation. During training, the model adjusts its parameters to minimize errors. Once trained, the model can generalize to new, unseen data, making predictions or classifications. Common types include supervised learning (with labeled data), unsupervised learning (finding patterns without labels), and reinforcement learning (learning through trial and error with rewards). ML applications range from image recognition to natural language processing, driving advancements in various industries. For example, the simplest training algorithm is the **perceptron learning algorithm**. It's a binary classification algorithm that learns a linear decision boundary to separate two classes. The steps are:

### Initialize Weights and Bias

Initialize weights (coefficients) and bias to small random values.

### For Each Training Example

Input features are multiplied by weights, and the bias is added.

If the result is greater than a threshold, predict one class; otherwise, predict the other.

### Update Weights

If the prediction is incorrect, adjust weights based on the error to improve future predictions.

### Repeat

Repeat the process with the updated weights until the model converges or a set number of iterations is reached.

The perceptron is a building block for more complex algorithms.

## ML Community Websites

MB Q: Is there a commonly known website for ML community?

ChatGPT A: Yes, there are several commonly known websites that serve as hubs for the machine learning (ML) community, providing resources, forums, and discussions. Some of the popular ones include:

### Stack Overflow - Machine Learning

Stack Overflow has a dedicated section for machine learning where practitioners and enthusiasts can ask and answer questions related to ML development, algorithms, and tools.

### GitHub

GitHub hosts numerous open-source ML projects and repositories. It's a valuable platform for collaboration, sharing code, and contributing to ML-related projects.

### Kaggle

Kaggle is a platform for data science and machine learning competitions. It also offers datasets, notebooks, and forums where the community collaborates on ML projects and challenges.

### Towards Data Science on Medium

Towards Data Science is a publication on Medium that features articles, tutorials, and discussions on various aspects of data science and machine learning.

### ArXiv - Machine Learning Section

ArXiv hosts a dedicated section for machine learning research papers. It's a primary resource for accessing and sharing the latest advancements in ML research.

### Reddit - r/MachineLearning

The r/MachineLearning subreddit on Reddit is a popular community for ML discussions, news, and Q&A. It covers a wide range of topics within the field.

### TensorFlow Forum

The TensorFlow community forum is a place where users discuss TensorFlow-related topics, seek help, and share knowledge about using the TensorFlow library.

### PyTorch Discussion Forum

The PyTorch discussion forum is the official platform for the PyTorch community to engage in discussions, share ideas, and seek help related to PyTorch.

These websites provide a mix of resources, discussions, and collaboration opportunities for individuals interested in machine learning, whether they are beginners or seasoned professionals.

## Professional Associations for QA testing AI / ML Generative Models

There aren't currently any established professional associations solely dedicated to QA testing of AI/ML generative models/products. However, the field is rapidly evolving, and several existing organizations are actively involved in promoting best practices and fostering collaboration in AI testing:

### International Software Testing Qualifications Board (ISTQB)

While ISTQB doesn't have a specific certification for AI testing yet, they offer a foundation level certification in "Agile Tester" that provides a strong base for understanding software testing principles applicable to AI models. Additionally, ISTQB is working on incorporating AI testing concepts into their future certifications. (<https://www.istqb.org/>)

### Association for Advancement of Artificial Intelligence (AAAI)

This organization holds conferences and workshops that cover topics related to AI testing and evaluation. You can find resources and connect with researchers and practitioners in the field through their website. (<https://aaai.org/>)

### IEEE Standards Association

The IEEE works on establishing standards for various technologies, including AI. While there aren't specific standards for AI testing yet, they might develop some in the future. You can stay updated on their website. (<https://www.ieee.org/>)

Here are some additional resources that might be helpful:

#### **Black Box Testing, White Box Testing and Grey Box Testing for Machine Learning Models:**

<https://www.nature.com/articles/s42256-019-0048-x> (This article explores different testing approaches for machine learning models.)

**A Survey on Testing Techniques for Deep Neural Networks:** <https://arxiv.org/pdf/2202.06372> (This research paper provides a comprehensive overview of testing techniques for deep learning models.)

While there's no single association dedicated to AI/ML generative model testing yet, the organizations and resources mentioned above offer valuable information and opportunities to stay informed and connected with the latest developments in this field.



## Guide to AI Tools

Claude Pro by Anthropic	*****	Same conversational feel as Claude 2. Pro version gives 5x the usage limit of the free version and priority access during peak capacity.	Processing large documents, organizing information, managing projects. Claude's ability to handle many file types makes it ideal for productivity tasks involving documents. Claude also has the largest context window of which can handle about 150,000 words.	95 countries
ChatGPT Plus (GPT-4) by OpenAI	*****	The most advanced LLM (and the best multi-modal model on our list) right now, but also charges \$20/month. Responses can often feel "robotic" or like reading a textbook. New model allows tools (Browse with Bing, Advanced Data Analysis, and DALL-E 3) to automatically be used without manual selection. GPTs make a great addition to the offering.	Research papers, essays, creative writing, content creation, and image generation. The added capability of plugins makes it suitable for more advanced use cases like code interpretation and data analysis.	Most countries
Claude 2 by Anthropic	*****	Using Claude feels more like having a natural conversation — making it seem more relatable and sometimes more professional. The go-to for basic use if you want quality but don't want to pay.	Processing large documents, organizing information, managing projects. Claude's ability to handle many file types makes it ideal for productivity tasks involving documents. Claude also has the largest context window of which can handle about 150,000 words.	95 countries
Perplexity AI	*****	Truly the gold standard in conversational search engines. Miles better than Bard and Bing Chat.	Research that requires the most up-to-date information with citations. Summaries of text, personalized answers based on interest and past search history. Tasks which require highly accurate and complete information.	Available globally
Poe by Quora	****	Access to multiple LLMs (even Meta's models and StableDiffusion image generation) is great but can be confusing for non-experts. Similar pricing to ChatGPT Plus at \$20/mo.	You want to access multiple LLMs from one platform. Desktop version is a nice feature.	No official list
ChatGPT (GPT-3.5) by OpenAI	****	Seems to hallucinate more than GPT-4, so use with caution and always fact check. The custom instructions feature can make up for pitfalls.	Brainstorming ideas, summarizing information, simple Q&A where some incorrect facts are acceptable. The speed makes it good for drafting and iterating quickly.	Most countries
Pi by Inflection	***	Has empathy but can be frustratingly slow. Use when you want a patient listener, not quick answers.	Having an empathetic conversation, exploring ideas, journaling, mental health applications where emotional intelligence is needed.	No official list
Bing Chat by Microsoft	***	The consistent citing of sources helps build trust in the information provided and image analysis is great. Can now generate images using DALL-E 3. An overall better option than Bard.	Getting quick answers to questions, research, and creative writing. Bing Chat's integration with web search and ability to provide annotated responses makes it ideal for on-the-fly research and fact checking.	Most countries where Microsoft does business, except for China.
Bard by Google	**	Only redeeming use cases are image analysis and extensions, otherwise don't waste your time until major improvements are made. New use cases are emerging like understanding and engagement with YouTube videos.	Recommending for image analysis and tasks where connections to Google Workspace, Google Flights, Google Maps, Google Hotels, and Youtube are useful.	Over 230 countries

## AI News

<https://www.therundown.ai/archive?page=3>